École Normale Supérieure
FIMFA — Statistique 2018–2019

Partiel

1. Les lois Gamma de paramètres $p > 0$ (forme) et $\lambda > 0$ (intensité) sont définies par leur densité sur $\mathbb{R}^+$

$$f_{p,\lambda}(x) = \frac{\lambda^p x^{p-1} \mathrm{e}^{-\lambda x}}{\Gamma(p)}$$

avec

$$\Gamma(p) = \int_0^\infty x^{p-1} \mathrm{e}^{-x} \mathrm{d}x \,.$$

Dans cet exercice, on considère le modèle formé par les lois Gamma et leur translatées où la loi de paramètre $\theta = (p, \lambda, \mu)^t$, $\mu \in \mathbb{R}$ est définie par sa densité sur $[\mu, \infty)$

$$f_{p,\lambda,\mu}(x) = \frac{\lambda^p (x-\mu)^{p-1} \mathrm{e}^{-\lambda(x-\mu)}}{\Gamma(p)} \,.$$

L'espace des paramètres est $\Theta =]0, \infty)^2 \times \mathbb{R}$.

(a) ($1\frac{1}{2}$ points) Le modèle est-il dominé ? exponentiel ? identifiable ?

(b) (3 points) Calculer l'espérance, la variance de la loi définie par $f_{p,\lambda,\mu}$. Calculer $\mathbb{E}_{p,\lambda,\mu}(X-\mu)^k$ pour k entier et $\kappa = \mathbb{E}_{p,\lambda,\mu}[(X - \mathbb{E}_{p,\lambda,\mu}X)^4] / \left(\mathrm{var}_{p,\lambda,\mu}(X)\right)^2$.

(c) Supposant μ connue, proposer un estimateur de p, λ à partir d'un n échantillon.
Préciser
  i. (1 point) si votre estimateur est sans biais,
  ii. (1 point) s'il définit une famille consistante,
  iii. (1 point) s'il définit une famille asymptotiquement normale (donner éventuellement la covariance asymptotique).

(d) Supposant $p = 1$ mais μ, λ inconnues,
  i. (2 points) Proposer un estimateur au maximum de vraisemblance $\widehat{\mu}_n, \widehat{\lambda}_n$ de μ, λ à partir d'un n échantillon. Est-il toujours défini ? unique ?
  ii. (2 points) Quelle est la loi de $\widehat{\mu}_n - \mu$?
  iii. (2 points) Quelle est la loi limite de $\sqrt{n}(\widehat{\lambda}_n - \lambda)$?
  iv. (1 point) Covariance de $\widehat{\mu}_n, \widehat{\lambda}_n$?
  v. (2 points) Proposer une région de confiance de taux de couverture asymptotique $1 - \alpha$ pour (μ, λ)

(e) Supposant maintenant $p = 2$, mais μ, λ toujours inconnues,
  i. (2 points) étudier l'existence, l'unicité de l'estimateur au maximum de vraisemblance.
  ii. (1 point) proposer un estimateur $(\widetilde{\mu}_n, \widetilde{\lambda}_n)$ construit à partir de la moyenne et de la variance empirique. Etudier sa consistance.
  iii. (1 point) Si vous pensez que l'estimateur de la question précédente est asymptotiquement normal préciser la matrice de covariance asymptotique.

(f) (1 point) On suppose maintenant que p, λ, μ sont inconnues. Proposer une famille d'estimateurs asymptotiquement normale pour (p, λ, μ) (il n'est pas demandé de calculer explicitement la covariance asymptotique).

2. Dans cet exercice, la loi du couple (X, Y) paramétrée par $(q, \theta, \sigma) \in \Theta =]0, 1[\times\mathbb{R}\times]0, \infty)$ est définie comme suit :

$$X \sim \text{Rademacher de paramètre } q$$
$$Y - \theta X \sim \mathcal{N}(0, \sigma^2) \text{ indépendante de } X$$

Une variable de Rademacher de paramètre q prend la valeur 1 avec probabilité q, la valeur -1 avec probabilité $1 - q$.

On observe $(X_i, Y_i)_{i\leq n}$, n tirages indépendants selon la loi (inconnue) de (X, Y).

(a) (1 point) S'agit d'un modèle dominé, exponentiel (en changeant évantuellement le paramétrage), identifiable ?

(b) (1 point) Peut-on définir un estimateur au maximum de vraisemblance dans ce modèle ?

(c) (2 points) Construire une région de confiance de niveau de couverture asymptotique $1 - \alpha$, pour (q, θ, σ).

(d) (2 points) Proposer un test de niveau $\alpha \in]0, 1[$ pour distinguer l'hypothèse nulle $\theta = 0$ contre l'alternative $\theta > 0$.

3. Dans cet exercice, $\theta_1, \ldots, \theta_d \in \mathbb{R}^d$ et $\sigma > 0$ sont inconnus. On observe pour chaque $i \leq d$, n_i tirages indépendants $X_{i,1}, \ldots, X_{i,n_i}$ selon $\mathcal{N}(\theta_i, \sigma^2)$. On notera $n = \sum_{i=1}^{d} n_i$, $\overline{X}_{n,i} = \sum_{j=1}^{n_j} X_{i,j}/n_i$.

(a) (2 points) Proposer un estimateur de σ^2_n de σ^2. Préciser la loi de cet estimateur. Proposer un intervalle de confiance de taux de couverture $1 - \alpha$ pour σ^2.

(b) (2 points) Proposer un estimateur $\hat{\theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_d)^T$, et pour chaque $i \leq d$, proposer un intervalle de confiance de taux de couverture $1 - \alpha$ pour θ_i

(c) (2 points) Proposer une région de confiance de taux de couverture $1 - \alpha$ pour le vecteur $(\theta_1, \ldots, \theta_d)^T$.

(d) (3 points) On veut maintenant tester l'hypothèse nulle $\theta_1 = \theta_2 \ldots = \theta_d$, contre l'alternative $\exists i, i' \leq d, \theta_i \neq \theta_{i'}$. Proposer un test de niveau α.

(e) (1½ points) Proposer un test de niveau α pour tester l'hypothèse nulle $\theta_1 = \theta_2$ contre l'alternative $\theta_1 > \theta_2$.

4. Dans cet exercice $(X, Y)^t \sim \mathcal{N}(\mu, K)$ avec

$$\mu = \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix} \qquad \text{et} \qquad K = \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix}$$

On observe $((X_i, Y_i)^t)_{i\leq n}$ n tirages indépendants selon $\mathcal{N}(\mu, K)$. On veut estimer le coefficient de corrélation ρ.

(a) On veut ajuster à ces données un modèle linéaire de la forme $Y = \alpha + \beta X + \eta Z$ où $\alpha, \beta \in \mathbb{R}$ $\eta > 0$, et Z est aléatoire, centrée, de variance 1 et indépendante de X.

i. (1 point) Si on veut minimiser le critère

$$\mathbb{E}\left[(Y - \alpha - \beta X)^2\right]$$

quel est le meilleur choix possible pour α, β ? (on suppose ici μ et K connus).

ii. (1 point) Si on choisit α, β comme dans la réponse à la question précédente, que peut-on dire de la la loi de Z ?

(b) Pour traiter un échantillon $((X_i, Y_i)_{i\leq n})$ collecté par tirages i.i.d. selon $\mathcal{N}(\mu, K)$, on ajuste une régression de Y par rapport à X par la méthode des moindre carrés. On note $\widehat{\alpha}, \widehat{\beta}, \widehat{\eta}$ les estimateurs de α, β, η. On note r le coefficient de corrélation empirique. On note $S^2 = \sum_{i=1}^{n}(Y_i - \widehat{\alpha} - \widehat{\beta}X_i)^2$ la somme des carrés des résidus et $S_x^2 = \sum_{i=1}^{n}(X_i - \overline{X}_n)^2$ (S_x^2/n est la variance empirique des X_i).

i. (1 point) Montrer que

$$\frac{\widehat{\beta}S_x}{S} = \frac{r}{\sqrt{1 - r^2}}$$

ii. (1 point) Quelle est la loi de $\sqrt{n-2}\frac{r}{\sqrt{1-r^2}}$ (conditionnellement à $X_1, \ldots, X_n$) lorsque $\rho = 0$?

iii. (1 point) Proposer un test pour distinguer l'hypothèse nulle $\rho = 0$ contre l'alternative $\rho > 0$.

Question :	1	2	3	4	Total
Points :	21½	6	10½	5	43
Score :					